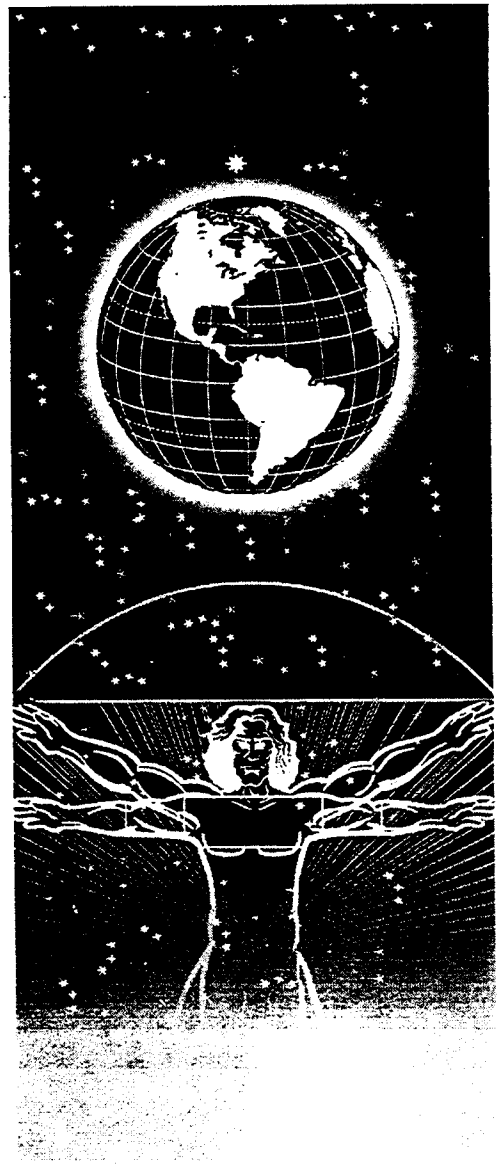
AL/HR-TP-1997-0010



UNITED STATES AIR FORCE
ARMSTRONG LABORATORY

USE OF RESPONSE LATENCIES TO ENHANCE SELF-REPORT PERSONALITY MEASURES

Frederick M. Siem

ARMSTRONG LABORATORY
Human Resources Directorate
Aircrew Training Research Division
Aircrew Performance Branch
7909 Lindbergh Drive
Brooks Air Force Base TX 78235-5352

October 1997

Approved for public release; distribution is unlimited.

AIR FORCE MATERIEL COMMAND
ARMSTRONG LABORATORY
HUMAN RESOURCES DIRECTORATE
AIRCREW TRAINING RESEARCH DIVISION
6001 South Power Road, Building 558
Mesa AZ 85206-0904

NOTICES

Publication of this paper does not constitute approval or disapproval of the ideas or findings. It is published in the interest of scientific and technical information (STINFO) exchange.

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Office of Public Affairs has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This paper has been reviewed and is approved for publication.

FREDERICK M. SIEM
Project Scientist

DEE H. ANDREWS
Technical Director

LYNN A. CARROLL, Colonel, USAF
Chief, Aircrew Training Research Division

Please notify AL/HRPP, 7909 Lindbergh Drive, Brooks AFB, TX 78235-5352, if your address changes, or if you no longer want to receive our technical reports. You may write or call the STINFO Office at DSN 240-3877 or commercial (210) 536-3877.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1997	Interim - March 1991 to January 1996

4. TITLE AND SUBTITLE

Use of Response Latencies to Enhance Self-Report Personality Measures

5. FUNDING NUMBERS

PE - 62205F
PR - 1123
TA - B1
WU - 01

6. AUTHOR(S)

Frederick M. Siem

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Armstrong Laboratory, Human Resources Directorate
Aircrew Training Research Division
Aircrew Performance Branch
7909 Lindbergh Drive
Brooks Air Force Base TX 78235-5352

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Armstrong Laboratory
Human Resources Directorate
Aircrew Training Research Division
6001 South Power Road, Building 558
Mesa AZ 85206-0904

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AL/HR-TP-1997-0010

11. SUPPLEMENTARY NOTES

Armstrong Laboratory Technical Monitor: Dr Frederick M. Siem
This paper was previously published in Military Psychology, Volume 8, Number 1 (1996), pp 15-27.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Previous research has demonstrated that latencies to personality inventory items explain unique variance in criteria such as peer ratings of personality (Popham & Holden, 1990). The present research was designed to determine if such latencies, reflecting a construct called the self-schema (Markus, 1977), would contribute to the prediction of pilot training performance. Latencies and scale scores from items on the Minnesota Multiphasic Personality Inventory (MMPI) were examined for a sample of U.S. Air Force pilot candidates. The results indicated that, as in previous studies, scale scores and response latencies tended to be correlated, although the pattern was not consistent across all trait dimensions. Furthermore, response latency measures for two trait dimensions added incremental validity over inventory scores alone to the prediction of flying training performance. The results were interpreted as providing support for further investigation of the utility of response latencies as indicants of the self-schema that may be useful for personnel selection.

14. SUBJECT TERMS

Measures; Personality; Personnel selection; Pilot performance; Pilot selection; Pilot training; Response latency; Self-schema; Traits

15. NUMBER OF PAGES
19

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

PREFACE

This research was conducted under Work Unit 1123-B1-01 (formerly 1123-A1-01), Pilot Selection and Classification Support, for the Armstrong Laboratory, Human Resources Directorate, Aircrew Training Research Division, Aircrew Performance Branch (AL/HRAA), located at Brooks Air Force Base, Texas. The laboratory principal investigator was Dr Frederick M. Siem.

I would like to express my appreciation to Joe Weeks, Malcolm Ree, Larry Stricker, and Lois Tetrick for their constructive comments, and to Cal Fresne, Paul Rioux, and Keith Weekley for their technical support.

This technical paper was published in *Military Psychology,* Volume 8, No. 1, pp 15-27.

USE OF RESPONSE LATENCIES TO ENHANCE SELF-REPORT PERSONALITY MEASURES

Previous research demonstrated that latencies to personality inventory items explain unique variance in criteria such as peer ratings of personality (Popham & Holden, 1990). This research was designed to determine if such latencies, reflecting a construct called the self-schema (Markus, 1977), would contribute to the prediction of pilot training performance. Latencies and scale scores from items on the Minnesota Multiphasic Personality Inventory (MMPI) were examined for a sample of U.S. Air Force pilot candidates. The results indicated that, as in previous studies, scale scores and response latencies tended to be correlated, although the pattern was not consistent across all trait dimensions. Furthermore, response latency measures for two-trait dimensions added incremental validity over inventory scores alone to the prediction of flying training performance. Results were interpreted as providing support for further investigation of the utility of response latencies as indicants of the self-schema that may be useful for personnel selection.

Personality can be defined broadly or narrowly. In the broad sense of the word, personality incorporates constructs such as traits, attitudes, moods, and even intelligence. In the narrow sense, particularly in the context of individual differences research, the term *personality is* used synonymously with the term *trait.* Beginning as early as Cicero, however, the term personality has been used to describe how a person appears to others (Allport, 1961), hence the derivation of the term from the Greek *persona,* the mask through which a stage actor spoke his lines. More recently, Hogan (1983) invoked the concept of personality as public reputation in his socioanalytic theory. Yet a third way in which the term personality has been used is to refer to "inferred, hypothesized, mediating internal states, structure, and organization of individuals" (Mischel, 1968, p. 4). This cognitively oriented definition of personality can be traced to concepts such as the self-concept as discussed by Cooley (1902) and Mead (1934).

Most research studies implicitly accept one of the narrow definitions of personality. To date, few studies have addressed distinctions between the different concepts of personality. One notable exception to this trend consists of a series of studies by Holden and colleagues (Holden & Fekken, 1987; Holden, Fekken, & Cotton, 1991; Popham & Holden, 1990). These studies compared both peer ratings of personality and inventory scores to measures of one particular aspect of the self-concept, the self-schema: "cognitive generalizations about the self, derived from past experience, that organize and guide the processing of self-related information" (Markus, 977, p. 64).

Evidence for the self-schema in the processing of information relevant to the self was found by a number of researchers (Markus & Wurf, 1987). For example, Kuiper (1981) found that the speed with which subjects made yes-no decisions about the self-descriptiveness of trait words demonstrated an "inverted-U RT [reaction time] effect" (p. 438). Decisions were made more quickly about adjectives either extremely like or unlike the self, compared to adjectives that were only moderately self-descriptive. A number of other studies also examined the speed with which self-referent decisions are made as constituting an index of the self-schema (Lewicki, 1984; Markus, 1977; Strube et al., 1986). Results of these studies were interpreted to indicate that the speed with which self-referent decisions are made is evidence of a well-articulated self-schema.

According to Holden and colleagues, the self-schema influences the time it takes a subject to respond to an item from a personality inventory as well as to simple trait words. In support of this notion, Holden and colleagues reported consistent findings that an individual who scores high on an inventory trait measure also manifests shorter latencies when endorsing items that describe the trait, as compared to an individual who scores low on the trait measure (Fekken & Holden, 1992; Holden & Fekken, 1987; Holden et al., 1991; Popham & Holden, 1990).

Before discussing these concepts further, it may be useful to define what is meant by endorsement and rejection of a personality item. Consider a bipolar trait dimension such as extroversion-introversion. An item can be stated in one of two directions, for example, "I am an extrovert" or "I am an introvert." If the items are scored or keyed for responses indicating extroversion, then endorsement results from agreeing with the first item and disagreeing with the second item. An item is considered rejected if the subject responds in opposition to how the item is keyed (i.e., disagreeing with the first item and agreeing with the second item). Endorsement and rejection, then, each necessitate joint consideration of how an item is scored and how a subject responds to the item.

An example may clarify how the self-schema is thought to affect the speed of endorsement and rejection of personality inventory items. Consider an individual who possesses a strong extroversion self-schema. Such a person will respond quickly to agree with a self-descriptive statement, such as "I am an extrovert," and also will be quick to disagree with statements that are not self-descriptive, such as "I am an introvert." In both cases, an individual with a strong self-schema manifests quick response times to endorsing an item in the keyed direction (i.e., agreeing with extroversion items and disagreeing with introversion items). Schematic individuals will manifest relatively longer response latencies when rejecting an item. That is, an extrovert will take relatively longer to disagree than to agree with the statement "I am an extrovert." Similarly, an extrovert will take relatively longer to agree than to disagree with the statement "I am an introvert."

The pattern of response latencies for individuals with low levels of a trait have been shown to be approximately the opposite of those for individuals with high levels of a trait (Popham & Holden, 1990). Individuals who report low levels of a particular trait are relatively quick to reject items and relatively slow to endorse items. That is, an introvert will take relatively longer to agree than to disagree with the statement "I am an extrovert." Similarly, an introvert will take relatively longer to disagree than to agree with the statement "I am an introvert."

Data from a number of samples (Fekken & Holden, 1992; Holden & Fekken, 1987) indicate that latency of item endorsement and inventory scores are correlated between -.25 and -.32 (i.e., the higher the scale score, the shorter the mean response time). Similarly, item rejection and scale scores are correlated at about the same magnitude but positively (the higher the scale score, the longer the time to reject an item). In addition, Popham and Holden (1990) found that item response time self-schema measures are correlated around -.20 with independent measures of

personality, such as roommate ratings of participant personality. Moreover, the latency measures predict unique variance in the roommate rating criteria over that predicted by inventory scores (Popham & Holden, 1990).

Referring back to the definitions of personality offered earlier, it can be argued that each of the measures used by Holden and colleagues (Fekken & Holden, 1992; Holden & Fekken, 1987; Popham & Holden, 1990) represents one aspect of personality: inventory scores corresponding to trait levels, inventory item response latencies representing the self-schema, and roommate ratings reflecting social reputation. If one accepts such arguments, then the research data indicate that these three aspects of personality are modestly correlated with each other, but that each to a large extent, also measures unique aspects of personality.

The studies by Holden and colleagues (Fekken & Holden, 1992; Holden & Fekken, 1987; Popham & Holden, 1990) employed samples from academic and clinical settings, and the purpose of the research was to demonstrate the relation of latency measures to measures such as roommate ratings and clinical evaluations. To date, little attention has been paid to examining the utility of schema-based response latency measures for personnel selection. In one of the few such studies published, Stricker and Alderton (1991) examined the utility of response latencies to biodata items for naval recruit selection. Their analyses indicated that latencies did not add unique validity to the prediction of training outcomes, but the latencies were useful for examining item characteristics.

This study was undertaken to assess the utility of inventory response latencies in the context of selecting military pilot candidates. Previous research demonstrated that personality characteristics are modestly correlated with performance in Air Force pilot training performance. Davis (1989), for example, reported a significant correlation of .13 between self-confidence and pilot training completion. Siem (1992) found a significant correlation of the same magnitude for self-confidence with an independent sample that was administered a different personality instrument. The Siem findings are of particular interest to this study because the instrument contained many of the same items used by Popham and Holden (1990), and response latencies were routinely collected. Thus, just as Popham and Holden found that response latencies to Minnesota Multiphasic Personality Inventory (MMPI) items added unique variance to the prediction of roommate rating criteria, it was hypothesized that latency measures would add unique variance to the prediction of pilot training performance.

METHOD

Participants

The full sample used in initial data analyses consisted of 509 student pilots entering into U.S. Air Force Undergraduate Pilot Training (UPT). A majority of the participants were men (99%; n = 503), and the average age was 23.8 years. The subsample included in the validity analyses consisted of 332 participants with complete data on the personality measures and training performance.

Measures

The Automated Aircrew Personality Profiler (AAPP) consisted of 202 items from several personality inventories that measured attributes thought by subject-matter experts to be important to success as an Air Force pilot. (See Siem, 1992, for details concerning the complete contents of the AAPP inventory.) The AAPP included 94 items from the original MMPI (Dahlstrom, Welsh, & Dahlstrom, 1972). Eighty-five of the MMPI items were used to generate scale scores based on the factor-analytically derived scoring scheme reported in Costa, Zonderman, Williams, and McCrae (1985). This scoring scheme was used rather than the one reported in Siem because Popham and Holden (1990) hypothesized that generating response latency measures of the self-schema requires internally consistent, content-based personality scales, as opposed to the original MMPI scales, which were heterogeneous in item content (Costa et al., 1985). Because only a subset of items from the MMPI was administered in this study, it was feasible to compute scores for only five of the nine scales developed by Costa et al. Scales were all scored in the direction of socially desirable characteristics; therefore, scale names were changed where appropriate. The neuroticism scale developed by Costa et al., for example, was labeled emotional stability, and psychoticism/infrequency was labeled communality/ frequency (or simply communality).[1] The number of items in each scale is shown in Table 1. The criterion was a dichotomous variable representing graduation-nongraduation (attrition) from UPT.

Table 1. MMPI Scale Scores From the AAPP

MMPI Scale Labels[a]	*Original No. of Items*	*AAPP Scale Labels*	*Reduced No. of Items*
Psychoticism/infrequency	120	Communality/frequency[b]	13
Neuroticism	65	Emotional stability	24
Extraversion	23	Extroversion	12
Inadequacy	30	Competency[b]	11
Cynicism	37	Trusting[b]	25
Total	275		85

Note. MMPI = Minnesota Multiphasic Personality Inventory; AAPP = Automated Aircrew Personality Profiler. [a]From Costa, Zonderman, Williams, and McCrae (1985). [b]AAPP scales scored in the direction of socially desirable characteristics.

Procedure

Prior to entry into flying training, the sample was administered the AAPP as part of a battery of experimental, computer-administered tests. For each MMPI item included in the AAPP, both

[1] One reason that the MMPI scales used in this research were shorter than the original scales was that items clearly designed for identifying signs of psychopathology ("Someone is trying to poison me") and deemed inappropriate for this application was eliminated. As a consequence, the psychoticism scale contained items mostly pertaining to the measurement of infrequently endorsed items, such as "It doesn't matter what becomes of me." This scale, when scored in the direction of socially desirable responses, thus became a measure of frequency of endorsement or what Gough (1987) refers to as communality.

endorsement responses (i.e., true-false) and response latencies (in milliseconds) were collected. Examinees were instructed to respond quickly to the personality items but were not explicitly instructed that response latencies were recorded. Each item was presented a single time, and examinees could not return to previous items.

The candidates took part in UPT, a 53-week course of instruction in flying subsonic and supersonic aircraft. Information on candidates' performance in UPT was routinely collected for operational purposes and archived at the Air Force Armstrong Laboratory Human Resources Directorate for research purposes.

Analysis

A procedure modeled after that used by Popham and Holden (1990) was used to create double-centered response latency measures. First, out-of-range latencies were defined as those that were less than 0.5 s or greater than 40 s. These latencies (120, or .1% of all latencies) were recoded to the minimum (0.5) or maximum (40) values, respectively. The resulting latencies were then standardized within subjects across items to control for confounding individual differences such as reading speed and motor speed. Next, response latencies were standardized by item across subjects to control for confounding item characteristics such as differences in item length and vocabulary level. Resulting standardized values greater than 3.0 (none were less than -3.0) were recoded to 3.0 (1,717, or 1.7% of all latencies).

In order to assess the reliability of the latency measures, split-half reliability coefficients were computed following a procedure described by Fekken and Holden (1992). For each participant, four scores were generated for each of the five scales, two representing mean response latency for endorsed items and two representing mean response latency for rejected items. Half the items from each scale were randomly assigned to one of the two half-scale response latency scores. Note that the number of items actually used to generate the mean varied from participant to participant as a function of the number of items endorsed and rejected for a particular scale for a particular individual. Split-half reliability coefficients were computed, and then the half-scale scores were combined into composite endorsed and rejected response latency measures for each of the five scales.

The validities of the scale and response latency measures were examined using correlational and regression analysis. For each trait, the predictive validity of each response latency measure was evaluated using hierarchical multiple regression. The predictiveness of a linear model with a scale score and a latency score was compared to the predictiveness of a linear model with scale scores alone. In addition, for each trait, the predictiveness of a linear model with both latency scores and the scale score was compared to the predictiveness of the scale score alone.

RESULTS

Means, standard deviations, and split-half reliability coefficients for the personality predictor measures are shown in Table 2. Reliability coefficients for the scale scores were all in the same range (.52 to .57), except for the communality items (.19). The latency measures demonstrated little evidence of reliability between the two half-scale measures (all *r*s < .20), although the coefficients for the endorsed item measures tended to be greater in magnitude than the corresponding rejected item measures.

Table 2. Descriptive Statistics for Predictor and Criterion Variables

Variable	*M*	*SD*	*Split-Half Reliability*
Predictor			
Scale score			
Communality	10.55	1.39	.19
Emotional stability	13.51	3.46	.57
Extroversion	8.37	2.45	.52
Competency	8.51	2.15	.57
Trusting	14.33	4.21	.57
Mean endorsed item latencies (RT_E)			
Communality	-.04	.30	.13
Emotional stability	-.03	.29	.11
Extroversion	-.06	.36	.12
Competency	-.03	.31	.14
Trusting	.00	.28	.13
Mean rejected item latencies (RT_R)			
Communality	.02	.64	.00
Emotional stability	.02	.30	.05
Extroversion	.16	.57	-.04
Competency	.00	.68	.05
Trusting	.04	.36	.15
Criterion			
UPT pass-fail	.80	.40	---

Note. N = 332. UPT = Undergraduate Pilot Training. UPT pass-fail is a dichotomous variable (1 = *pass*, 0 = *fail*). Higher scale scores indicate more of the indicated trait. Mean latencies are based on standardized item latencies. A dash indicates that data are not available.

The correlations of UPT pass-fail with the scale scores, mean response latency for endorsed items, and mean response latency for rejected items are shown in Table 3. The data reported in Table 3 indicate that the magnitudes of the relations between personality scale scores and UPT outcome were fairly small, although scores from two of the scales (communality and trusting) were statistically significant. Two of the response latency measures (extroversion endorsed items and trusting rejected items) were also significantly correlated with UPT outcomes.

The intercorrelations of the personality latency measures and scale scores are shown in Table 4. As was expected from previous research (e.g., Popham & Holden, 1990), scale scores tended to be correlated negatively with latencies for endorsed items and positively with latencies for rejected items. The pattern for the communality dimension was the weakest. Communality was not significantly correlated with either corresponding latency measures. However, endorsed and rejected latencies for emotional stability items were significantly correlated with the communality scale score. The emotional stability scale score was significantly correlated with both response latency measures to the trusting items, and the trusting scale score was correlated with rejected latency for emotional stability items.

Table 3. Correlations with UPT Pass-Fail for Five Personality Scales

	UPT Pass-Fail Correlation With:		
Trait	*Scale*	RT_E	RT_R
Communality	.14*	-.05	-.09
Emotional stability	.07	-.05	.02
Extroversion	.06	-.12*	.00
Competency	.10	-.03	.04
Trusting	.14*	.04	.12*

Note. N = 332. UPT = Undergraduate Pilot Training; RT_E = mean latency for endorsed items; RT_E = mean latency for rejected items.
*$p < .05$.

Table 4. Intercorrelations of Three Types of Variables for Five Personality Scales

		Scale				
		Communality	*Emotional Stability*	*Extroversion*	*Competency*	*Trusting*
RT_E						
	Communality	-.07	-.03	-.01	-.03	-.09
	Emotional stability	-.20***	-.16**	-.01	-.12	-.06
	Extroversion	-.02	.02	-.29***	-.13*	.03
	Competency	-.11*	-.19***	-.10	-.35***	-.14*
	Trusting	-.17**	-.25***	-.05	-.14*	-.37***
RT_R						
	Communality	-.04	.00	.02	.07	-.08
	Emotional stability	.15**	.24***	-.04	.08	.22***
	Extroversion	-.02	-.06	.14*	.07	-.01
	Competency	.06	-.05	.09	. 08	.00
	Trusting	.10	.22***	.08	.10	.23***

Note. N = 332. RT_E = mean latency for endorsed items; RT_E = mean latency for rejected items.
*$p < .05$. **p , $< .01$. ***$p < .001$.

The intercorrelations of the endorsed and rejected latency measures are shown in Table 5. Most of the correlations between corresponding endorsed and rejected item latencies for the same trait were negative, as was expected. However, the correlations were significant for only two of the trait dimensions—emotional stability and trusting.

The incremental validity of the latency scores compared to personality scale scores was assessed through a series of hierarchical multiple regressions following the procedure used by Popham and Holden (1990). For each personality dimension, two regressions were conducted. The criterion for both regressions was UPT outcome. Predictor scores were entered in two steps. On the first step, the scale score was entered into a prediction equation. For one regression, only the mean latency score for endorsed items was entered on the second step. For the second regression, only the mean latency score for rejected items was added to the prediction equation on the second step. For each regression, the change in the magnitude of the relation between the criterion and the predictor set was evaluated at each step of the process.

The results of the multiple regression analyses are summarized in the middle of Table 6. As those data indicate, only the response latency measure for endorsed extroversion items demonstrated validity incremental to that of the scale score alone. Examination of mean latency scores indicated that UPT graduates were quicker to endorse extroversion items (M = -.06) than were attrites (M = .03).

Table 5. Intercorrelation of Endorsed Item and Rejected Item Latency Measures

	Rejected Item Latency Measure				
Endorsed Item Latency Measure	*Communality*	*Emotional Stability*	*Extroversion*	*Competency*	*Trusting*
Communality	-.03	-.05	-.05	-.04	-.20***
Emotional stability	.05	-.20***	.07	.03	.17**
Extroversion	.00	-.09	-.02	-.04	-.01
Competency	-.06	-.11*	-.10	-.07	-.11*
Trusting	-.09	-.14**	-.01	-.06	-.25***

Note. $N = 332$.

*p < .05. **p < .01. ***p < .001.

Somewhat different results emerged from a third series of regressions. For these regressions, the incremental validity of both latency measures was assessed relative to the validity of the scale score alone. The results of these regressions are shown on the right of Table 6. These analyses indicated that the combination of the latency measures for trusting added significantly to the validity of the scale score.

Table 6. Validity of Personality Scale Scores and Incremental Validity of Latency Measures for Prediction of UPT Performance

		Predictor Variables					
		$+RT_E$		$+RT_R$		$+(RT_E + RT_R)$	
Trait	*r*	*R*	*ΔR*	*R*	*ΔR*	*R*	*ΔR*
Communality	.137*	.142*	.005	.160*	.023	.165*	.028
Emotional stability	.069	.080	.011	.069	.000	.080	.011
Extroversion	.056	.120	.064*	.056	.000	.120	.064
Competency	.095	.095	.000	.101	.006	.102	.007
Trusting	.138*	.167**	.029	.162**	.024	.197**	.030*

Note. $N = 332$. UPT = Undergraduate Pilot Training; RT_E = mean latency for endorsed items; RT_R = mean latency for rejected items. Incremental validity of individual latency measures is described in Columns 3 to 6. Incremental validity for both latency measures in combination is described in Columns 7 to 8.

*$p < .05$. **$p < .01$.

DISCUSSION

The results of this study provide some insights into the potential benefits and problems associated with the use of response latencies to personality items for personnel selection. Evidence for the potential benefits of response latency measures was the observation that some of the self-schema measures used in this study added incremental validity to personality scale scores for the prediction of training performance. One problem with those measures, however, was the observed low split-half reliability coefficients, suggesting that the instrument used in this study may not have been optimal for assessing the value of response latencies as a measure of the self-schema construct.

Evaluation of the low reliabilities for the self-schema measures used in this study requires consideration of both the instrument used to generate latencies and the procedures used to generate reliability coefficients. With regard to the latter, Fekken and Holden (1992) noted the limitations of examining split-half reliability coefficients based on separate examination of endorsed and rejected items from scales with relatively few items to begin with, which results in latency scores based on even fewer items. Fekken and Holden used a test-retest procedure to assess the reliability of their self-schema measures and found that 13 of 20 endorsed item latency measures and 9 of the rejected item latency measures were significantly correlated ($rs >$ .26) between the first and second administrations of the instrument. It should also be noted, however, that the mean reliabilities for endorsed and rejected item latency measures were .42 and .26, respectively, as compared to .91 for the scale scores. Regardless of procedure or inventory used, the available evidence suggests that personality items produce latency-based self-schema scores that are less reliable than their corresponding scale scores.

With regard to the instrument used in this study, it should be noted that the items were not designed specifically for the collection of item response latencies to measure the self-schema. To date, all research studies along these lines have used available instruments that were designed to collect reliable, valid inventory responses rather than response latencies. Traditional personality instruments are developed based only on items that produce reliable and valid trait scores, but it does not necessarily follow that such items will also produce reliable and valid self-schema measures. It may well be that items designed to elicit reliable response latencies may differ in some aspects (e.g., number of words or lexical complexity) from items that produce reliable responses. One challenge for future research, therefore, is to develop instruments explicitly designed to measure the self-schema using latency scores rather than capitalizing on available personality instruments that may or may not be best suited for such a purpose.

Another limitation of this study was that it relied on only one type of criterion measure—overall performance in flight training. Preferably, other criterion measures would have been collected similar to those used by Popham and Holden (1990), such as peer ratings of personality. In addition, it would be desirable in future research to collect data comparable to those from this study in personnel situations where stronger relations have been identified between personality measures and performance criteria.

Another issue for consideration in future research is the extent to which latency measures can be intentionally manipulated. In this study, data were collected for research purposes only; measurement of latencies was unobtrusive, and the collection of such latencies was not made explicit to the participants. Similar unobtrusive collection of such latency measures for operational use in a large-scale military organization is more problematic because such information quickly finds its way into commercial publications that instruct applicants how to improve their test scores (e.g., Wiener, 1989). Hence, before latency-based, self-schema measures are used for personnel selection, it would be worthwhile to determine how examinee awareness of latency measurement affects the subsequent validity of such measures.

A final issue for future research is a reexamination of the procedures used to generate latency-based, self-schema measures with a focus on psychometric and scientific concerns. From a scientific perspective, it would be valuable to know how much variance in item latency scores is explained by item factors, such as number of words, and by individual differences, such as cognitive ability, but the standardization of response latencies precludes examination of such issues. From a psychometric perspective, the double-standardization procedure also warrants further evaluation. For example, double-standardization procedures can result in different scores depending on the order in which the procedure is conducted (within-item then within-subject compared to within-subject then within-item). Furthermore, standardization procedures typically have been conducted in a sample-dependent fashion, with latencies standardized on the basis of observed data for a particular sample, a methodology that would be impractical for

operational administration of such an instrument to individual job applicants. Future research might benefit from examining possible alternatives to the double-standardization procedure, such as adjusting item latencies based on a normative sample and adjusting individual latencies based on a measure of simple response time collected in conjunction with the administration of the personality inventory.

In conclusion, much remains to be learned about the cognitive processes that underlie responding to items from personality inventories. This research produced some evidence that the self-schema contributes to such responses, but the results also suggest a number of directions for additional research. It can be expected that such research will add to understanding of the type of measures examined in this study and will help to clarify their potential use in personnel selection applications.

ACKNOWLEDGMENTS

The views expressed in this article are those of the author and do not necessarily reflect official views of the Air Force or the Department of Defense.

Portions of this article were presented at the meeting of the Military Testing Association, San Antonio, TX, November 1991 and at the meeting of the American Psychological Association, Washington, DC, August 1992. I thank Joe Weeks, Malcolm Ree, Larry Stricker, Lois Tetrick, and three anonymous reviewers for their helpful comments. I also gratefully acknowledge the technical support provided by Calvin Fresne, Paul Rioux, and Keith Weekley.

REFERENCES

Allport, C.W. (1961). *Pattern and growth in personality.* New York: Holt, Rinehart & Winston.

Cooley, C.H. (1902). *Human nature and the social order.* New York: Scribner's.

Costa, P.T., Zonderman, A.B., Williams, R.B., & McCrae, R.R. (1985). Content and comprehensiveness in the MMPI: An item factor analysis in a normal adult sample. *Journal of Personality and Social Psychology,* 48, 925-933.

Dahlstrom, W.G., Welsh, G.S., & Dahlstrom, L.E. (1972). *An MMPI handbook* (Vols. 1-2). Minneapolis: University of Minnesota Press.

Davis, R.A. (1989). *Personality: Its use in selecting candidates for Air Force pilot training* (Rep. No. AU-ARI-88-8). Maxwell Air Force Base, AL: Air University Press.

Fekken, G.C., & Holden, R.R. (1992). Response latency evidence for viewing personality dimensions as schema. *Journal of Research in Personality,* 26, 103-120.

Gough, H. (1987). *California Psychological Inventory administrator's guide.* Palo Alto, CA: Consulting Psychologists Press.

Hogan, R. (1983). A socioanalytic theory of personality. In M. M. Page & R. A. Dienstbier (Eds.), *Nebraska symposium on motivation: Vol. 29. Personality—Current theory and research* (pp. 55-89). Lincoln: University of Nebraska Press.

Holden, R.R., & Fekken, G.C. (1987, August). *Reaction time and self-report psychopathological assessment: Convergent and discriminant validity.* Paper presented at the American Psychological Association Convention, New York.

Holden, R.R., Fekken, G.C., & Cotton, D.H.G. (1991). Assessing psychopathology using structured test item response latencies. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3, 111-118.*

Kuiper, N.A. (1981). Convergent evidence for the self as prototype: The "inverted-U RT effect" for self and other judgments. *Personality and Social Psychology Bulletin,* 7, 438-443.

Lewicki, P. (1984). Self-schema and social information processing. *Journal of Personality and Social Psychology,* 47, 1177- 1190.

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology, 35,* 63-78.

Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology,* 38, 299-337.

Mead, G.H. (1934). *Mind, self, and society.* Chicago: University of Chicago Press.

Mischel, W. (1968). *Personality and assessment.* New York: Wiley.

Popham, S.M., & Holden, R.R. (1990). Assessing MMPI constructs through the measurement of response latencies. *Journal of Personality Assessment,* 54, 469-478.

Siem, F. M. (1992). Predictive validity of an automated personality inventory for Air Force pilot selection. *The International Journal of Aviation Psychology, 2,* 261-270.

Stricker, L.J., & Alderton, D.L. (1991). *Response latency measures for biographical inventories.* (Rep. No. RR-91-19-ONR). Princeton, NJ: Educational Testing Service.

Strube, M. J., Berry, J.M., Lott, C.L., Fogelman, R., Steinhart, G., Moergen, S., & Davison, L. (1986). Self-schematic representation of the Type A and B behavior patterns. *Journal of Personality and Social Psychology, 51,* 170- 180.

Wiener, S. (1989). *Military flight aptitude tests.* New York: Prentice-Hall.